# SCREENING VOICE DISORDERS WITH THE GLOTTAL TO NOISE EXCITATION RATIO

*N. Sáenz-Lechón, V. Osma-Ruiz, R. Fraile, J. I. Godino-Llorente, P. Gómez-Vilda*
Universidad Politécnica de Madrid, Spain

## ABSTRACT

This work evaluates the capabilities of the Glottal to Noise Excitation Ratio for the screening of voice disorders. A lot of effort has been made using this parameter to evaluate voice quality, but there do not exist studies that evaluate the discrimination capabilities of this acoustic parameter to classify between normal and pathological voices. A set of 226 speakers (53 normal and 173 pathological) taken from a voice disorders database were used to evaluate the usefulness of this parameter for discriminating normal and pathological voices. In order to evaluate this parameter, the effect of the bandwidth of the Hilbert envelopes and the frequency shift have been analyzed, concluding that a good discrimination is obtained with a bandwidth of 1000 Hz and a frequency shift of 300 Hz. The results confirm that the Glottal to Noise Excitation Ratio provides reliable measurements in terms of discrimination among normal and pathological voices, comparable to other classical long-term noise measurements found in the literature, such as Normalized Noise Energy or Harmonics to Noise Ratio, so this parameter is a good candidate to be used for screening purposes.

## 1. INTRODUCTION

A wide range of complementary acoustic parameters have been developed to measure different irregularities or perturbations. Most of the parameters [1-5] found in existing literature focus on perturbation measurements and the evaluation of voice pathologies by means of long-term parameters calculated by averaging local short-time perturbations measured from the speech. These parameters are usually grouped into three main categories: a) amplitude perturbation (or shimmer parameters); b) frequency perturbation (or jitter parameters); and, c) noise parameters.

Noise parameters give an indication of the noise content of the signal and have an extensive application in the evaluation of voice quality (because of their relationship with many dysphonia [6;7], and because they are well correlated with the perceptual ratings) and for screening purposes.

Apart from the physical interpretation of each acoustic parameter and its ability to measure different aspects of voice quality, it is very important to know how likely it is that a voice register is normal or not given each of the previously mentioned acoustic parameters. In the context of screening, there does not exist a single feature which is capable of differentiating between normal and pathological voices, since voice pathologies tend to combine different kinds of perturbations. To date, very few studies have evaluated the discriminative capabilities of the acoustic parameters. With the same database utilized in this study, Parsa evaluated the discriminative capabilities of several noise features [8]: NNE, SNR, zHNR, fHNR, PA, SPR; reporting accuracies equal to 79.8, 82.5, 83.3, 88.6, 94.7 and 98.7% respectively. Yumoto [3] first proposed the HNR parameter for acoustic discrimination of voice disorders reporting an error rate of 16.7%. Kasuya [1] obtained a classification accuracy of 78.6% for the NNE and 74.1% for the HNR. Other works [6;8-11] indicate that an accurate screening can be carried out using a combination of several of the above mentioned acoustic parameters, enabling each individual voice utterance to be quantified by a single set of one-dimensional parameters (similar to those enumerated above). Although the multidimensional studies reported a good efficiency in screening (obtaining accuracy of up to 96%) [10;11], such analysis is not always easy to interpret from the perspective of a human evaluator and it is usually carried out by methods based on complex pattern recognition techniques. In this sense, the best acoustic features for screening would be those with the lowest correlation against the others and with the best discrimination capabilities. Among these lines, the GNE, as reported in [5;12], provides a low correlation with respect to the amplitude perturbation and noise features, but the classification accuracy of the GNE has never been documented consistently. Moreover, an advantage of this parameter is that its calculation is not based on a previous estimation of the fundamental frequency, a difficult task in the presence of pathology. Thus, the GNE is a good candidate to be used for screening purposes.

The main goal of this work is to evaluate the discrimination capabilities to separate normal and pathological voices (i.e. screening accuracy) of the GNE, and comparing its reliability with several noise and amplitude/frequency perturbation parameters found in related literature. To date, the GNE has only been used previously in voice quality studies to represent the "hoarseness diagram" [12], but there are no existing extensive studies about the validity of this particular parameter for screening purposes in related literature.

## 2. THE GLOTTAL TO NOISE EXCITATION RATIO

The algorithm to calculate the Glottal to Noise Excitation Ratio (GNE) was first proposed by Michaelis in [5;12]. The GNE represents an interesting approach to quantify the amount of excitation due to vocal fold oscillations versus the excitation given by turbulent noise. Thus, it is closely related to breathiness, and it is considered a reliable measure for the relative noise level even in the presence of strong amplitude and frequency perturbations.

In contrast to other acoustic parameters -such as jitter, shimmer, NNE and NHR-, one of the most interesting aspects of the GNE is that its calculation can be considered very robust because it does not require a previous estimation of the fundamental frequency, a difficult task in the presence of pathology, and it is thus suited even to highly irregular glottal oscillations.

The GNE is based on the correlation between Hilbert envelopes of different frequency channels uniformly distributed throughout the spectrum (channels with a bandwidth: BW; and separated with a frequency shift: FShift). Triggered by each glottis closure, all the frequency channels are simultaneously excited so that the Hilbert envelopes in all channels should be of the same shape, leading to high correlation between envelopes. In the case of turbulent signals (whisper or pathological voices) a narrow band noise is excited in each frequency channel. These narrow band noises can be considered uncorrelated if the windows that define the frequency bands do not overlap too much [5]. The noise leads to a lower correlation between envelopes. The output of each frequency channel can be considered as a pass-band signal, and the Hilbert envelope an estimation of the amplitude envelope of these signals.

## 3. DATABASE

The tests have been carried out using a commercially available database developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Labs (MEEI). It was compiled partly at the MEEI Voice and Speech Lab. and partly at Kay Elemetrics Corp. and it features recordings of the sustained phonation of vowel /ah/ (53 normal and 657 pathological files).

The voices were recorded with a sampling frequency of 50 kHz and 16 bits of quantization. Each subject was asked to produce a sustained phonation at a comfortable pitch and loudness. The process was repeated three times for each subject, and a speech pathologist chose the best sample for the database. The acoustic samples are sustained phonations (1~3 s. long) of vowel /ah/ from patients (males and females) with normal voices and a wide variety of organic, neurological, traumatic, and psychogenic voice disorders. The samples were edited to remove the first and the last part of the utterance to avoid onset and offset effects. The duration of the pathological records stored in the database is around 0.8 s., whereas the normal voices are 3 s. long. When necessary, a downsampling with a previous half band filtering has been done to adjust every utterance to the sampling rate of 25 kHz.

The database has been segmented according to the criteria explained in [8] and previously used in other studies [13]. These criteria ensure that all the files are labelled with their diagnosis, and also that gender and age are uniformly distributed amongst the samples belonging to both classes. The normal talkers exhibited no abnormal vocal characteristics and had no history of voice disorders. The final subset taken from the database contains 53 normal and 173 pathological talkers.

## 4. RESULTS

The accuracy of the parameters is tested and compared using the Relative Operating Characteristic (ROC) [14] plots. The ROC curve displays the diagnostic accuracy expressed in terms of sensitivity (or true positive rate) against 1-specificity (or false acceptance rate) at all possible decision threshold values. The ROC is analyzed calculating the Area Under the Curve (AUC) and its standard error (SE) as suggested in [14]. The AUC represents an estimation of the screening accuracy, and the SE provides the confidence intervals of this measurement.

The discriminative capabilities of the GNE are presented graphically in Figure 1, Figure 2, and Figure 3. Figure 1 depicts the AUC among its standard error SE for the experiments performed. Figure 2 show the influence of the FShift on the screening accuracy through three plots corresponding to the GNE extracted with 60 ms windows and BW=1000 Hz for a FShift equal to 100, 200 and 300 Hz.

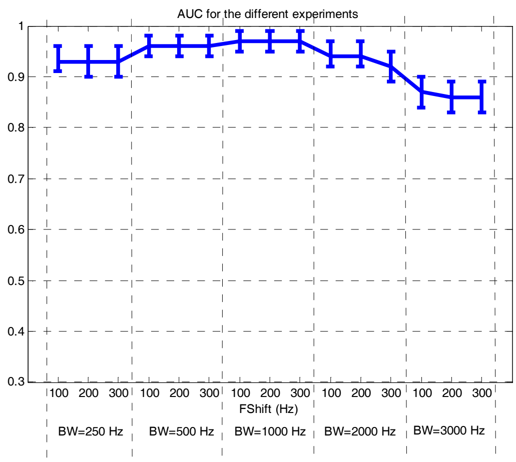


Figure 1: Screening accuracy of the $GNE_l$ parameter extracted with 60 ms windows for the 15 experiments performed: BW={250, 500, 1000, 2000, 3000} Hz, FShift={100, 200, 300} Hz.

Figure 3 shows the influence of the BW on the screening accuracy. This plot shows the curves that correspond to the GNE extracted with 60 ms windows and FShift=300 Hz for a BW equal to 250, 500, 1000, 2000 and 3000 Hz. The plots show that if BW≥1000 Hz, the lower the BW the better the discrimination capability of the parameter is; and also if BW≤500 Hz, the lower the BW the worse the discrimination is. In addition, Figure 3 shows that the efficiency in the detection of voice pathologies reached 90% using BW=1000 Hz and FShift=300 Hz. The worst results in terms of discrimination were obtained with the smallest and largest bandwidths (BW=250 Hz and BW=3000 Hz) tested.

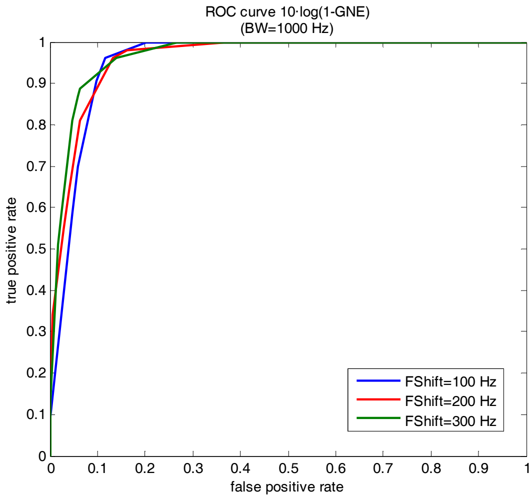


Figure 2: Accuracy of the GNE parameter. The bandwidth used is 1000 Hz, and the frequency shift (FShift) 100, 200 and 300 Hz. a) ROC plots for the different frequency shifts: FShift=100 Hz (AUC=0.97), FShift=200 Hz (AUC=0.97); and FShift=300 Hz (AUC=0.97).

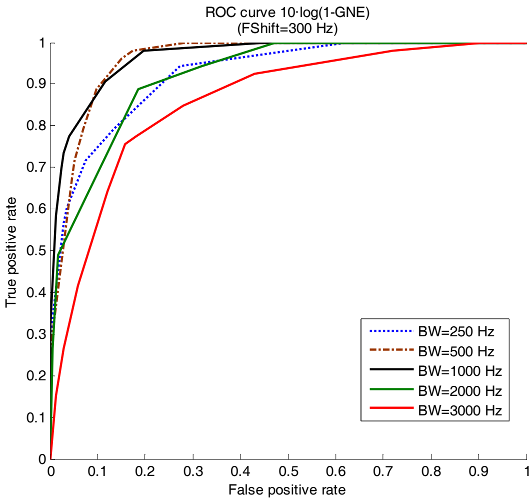


Figure 3: Accuracy of the GNE parameter. The frequency shift is 300 Hz, and the bandwidth (BW) 250, 500, 1000, 2000 and 3000 Hz. a) ROC plots for the different bandwidths: BW=250 Hz (AUC=0.93), BW=500 Hz (AUC=0.96), BW=1000 Hz (AUC=0.97), BW=2000 Hz (AUC=0.92); and BW=3000 Hz (AUC=0.86).

For the shake of comparison, the discriminative capability of several noise and perturbation parameters has also been evaluated and compared with the GNE calculated with a BW=1000 Hz and FShift=300 Hz. The discriminative capability evaluated in terms of the efficiency in the detection of voice disorders represents an estimation of the percentage of voices that have a deviation in the phenomenon represented by the acoustic parameter.

Figure 4 shows the plots of the ROC curves for the GNE compared with other noise parameters existing in the related literature, such as: NNE, CHNR, HNR and VTI. Figure 4 depicts graphically the comparison of the different noise parameters showing that the GNE (under the proposed configuration: BW=1000 Hz and FShift=300Hz) provided similar results to the other noise features. In any case, each noise parameter measures different aspects of the phenomenon, so their measurements should be considered complementary.

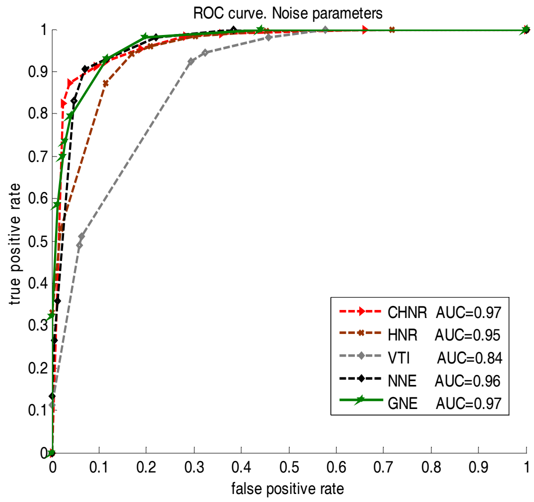


Figure 4: ROC plot to show and compare the ability to discriminate among normal and pathological voices

## 5. CONCLUSIONS

To date, and among the noise parameters, there are not many studies that evaluate the diagnostic capabilities of the GNE. However, this parameter showed a significant potential for screening since the GNE represents an interesting approach to quantify the amount of voice excitation by vocal fold oscillations versus excitation by turbulent noise.

It is acknowledged that the Hilbert envelopes provide accurate instantaneous values only for pure tones. The wider the BW, the more harmonics would be included in it deeming the envelope less accurate. In contrast, the narrower the BW, the bigger the chance that for higher fundamental frequencies some filter channels may not include harmonic energy, which would affect the GNE values with regard to the detection of pathology. Thus, the optimal BW should be the balance, which minimizes these two extreme effects. The results demonstrated that the performance starts to decline at BW≤500 Hz, because the method needs to warrant at least two or three harmonics in each filter bank; and, on the other hand, the performance declines at BW>1000 Hz, because -as commented above- the estimation of the Hibert envelope is less accurate. In order to decrease the computational requirements and to ensure at least three harmonics in each filter bank a good choice is to use BW=1000 Hz. On the other hand, there is no evidence that the FShift influences the discrimination capabilities of the parameter; furthermore, once again, in order to decrease the computational load of the algorithm a good trade-off is to use FShift=300 Hz.

These findings appear to contradict the methodology proposed in [12], where the GNE is calculated using a BW=3000 Hz to evaluate voice quality by means of the "hoarseness diagram". In [12] the authors demonstrated by means of correlation that the GNE with BW=3000 Hz correlates less with jitter and shimmer measurements, so they concluded that using this BW the parameter is less affected by frequency or amplitude perturbation. However, both results can be considered complementary, and the choice of the bandwidth depends on the use given to the parameter: whether the GNE is used to build the "hoarseness diagram" the best configuration is that reported in [12] reducing the correlation with the periodicity perturbation parameters; but if the GNE is used for screening purposes, the results demonstrate that a smaller bandwidth is preferred (BW=1000 Hz), although the correlation with the periodicity perturbation parameters increases (i.e. the GNE is integrating noise and aperiodicity information).

On the other hand, a complete analysis of voice requires a multidimensional analysis. However the single features with a clear physical interpretation still remain appropriate for the evaluation and screening of voice disorders in the clinical environment. In this sense, the set of noise and periodicity perturbation parameters calculated has a clear utility for the discrimination between normal and pathological voices, with classification rates over 73% for each parameter alone. Regarding the GNE, the efficiency for screening reached 90% (with BW=1000 Hz and FShift=300), comparable to other noise measurements such as CHNR and NNE but with the advantage of not requiring a previous estimation of the fundamental frequency. In general terms, the results suggest that the noise features (and the GNE among them) are good indicators of the presence or absence of pathology, whereas the efficiency of the periodicity perturbation parameters for screening purposes is lower. This fact does not mean that the periodicity perturbation parameters perform better or worse than the noise parameters; it simply means that these results have to be understood on the basis that the noise perturbations are more frequent in the presence of disorders than the periodicity perturbations.

## 6. ACKNOWLEDGEMENTS

This work was carried out under grant TEC2006-12887-C02 from the Ministry of Science and Technology of Spain.

## 7. REFERENCES

[1] Kasuya, H., Ogawa, S., Mashima, K., and Ebihara, S., "Normalized noise energy as an acoustic measure to evaluate pathologic voice," Journal of the Acoustical Society of America, vol. 80, no. 5, pp. 1329-1334, Nov.1986.

[2] de Krom, G., "A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals," Journal of Speech and Hearing Research, vol. 36, no. 2, pp. 254-266, Apr.1993.

[3] Yumoto, E., Gould, W. J., and Baer, T., "Harmonics-to-noise ratio as an index of the degree of hoarseness," Journal of the Acoustical Society of America, vol. 71, no. 6, pp. 1544-1550, 1982.

[4] Feijoo, S. and Hernández-Espinosa, C., "Short-term stability measures for the evaluation of vocal quality," Journal of Speech and Hearing Research, vol. 33, pp. 324-334, June1990.

[5] Michaelis, D., Gramss, T., and Strube, H. W., "Glottal-to-Noise Excitation ratio - a new measure for describing pathological voices," Acustica/Acta acustica, vol. 83, pp. 700-706, 1997.

[6] Yumoto, E., Sasaki, Y., and Okamura, H., "Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness," Journal of Speech and Hearing Research, vol. 27, no. 1, pp. 2-6, Mar.1984.

[7] Yanagihara, N., "Significance of harmonic changes and noise components in hoarseness," Journal of Speech and Hearing Research, vol. 10, pp. 531-541, 1967.

[8] Parsa, V. and Jamieson, D. G., "Identification of pathological voices using glottal noise measures," Journal of Speech, Language and Hearing Research, vol. 43, no. 2, pp. 469-485, Apr.2000.

[9] Hadjitodorov, S., Boyanov, B., and Teston, B., "Laryngeal pathology detection by means of class-specific neural maps," IEEE Transactions on Information Technology in Biomedicine, vol. 4, no. 1, pp. 68-73, Mar.2000.

[10] Hadjitodorov, S. and Mitev, P., "A computer system for acoustic analysis of pathological voices and laryngeal disease screening," Medical Engineering & Physics, vol. 24, no. 6, pp. 419-429, July2002.

[11] Ritchings, R. T., McGillion, M. A., and Moore, C. J., "Pathological voice quality assessment using artificial neural networks," Medical Engineering & Physics, vol. 24, no. 8, pp. 561-564, 2002.

[12] Michaelis, D., Fröhlich, M., and Strube, H. W., "Selection and combination of acoustic features for the description of pathologic voices," Journal of the Acoustical Society of America, vol. 103, pp. 1628-1639, 1998.

[13] Godino-Llorente, J. I., Gómez-Vilda, P., and Blanco-Velasco, M., "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters," IEEE Transactions on Biomedical Engineering, vol. 53, no. 3, pp. 1943-1953, 2006.

[14] Hanley, J. A. and McNeil, B. J., "The meaning and use of the area under a receiver operating characteristic (ROC) curve," Radiology, vol. 143, no. 1, pp. 29-36, Apr.1982.